(12) UK Patent Application (19) GB (11) 2 085 629 A

(54) Object recognition

(57) An object recognition system includes several pivotable cameras 2—5 and frame stores 6—9 for capturing images of a moving object from various aspects. The data is passed to compressors 10—13 to provide simplified correlation of data in correlators 14—17 also receiving data from a compressed image library 18 to allow the object and/or its orientation to be determined. A single camera may alternatively be used, to provide different views of the object.

Fig.1

GB 2 085 629 A
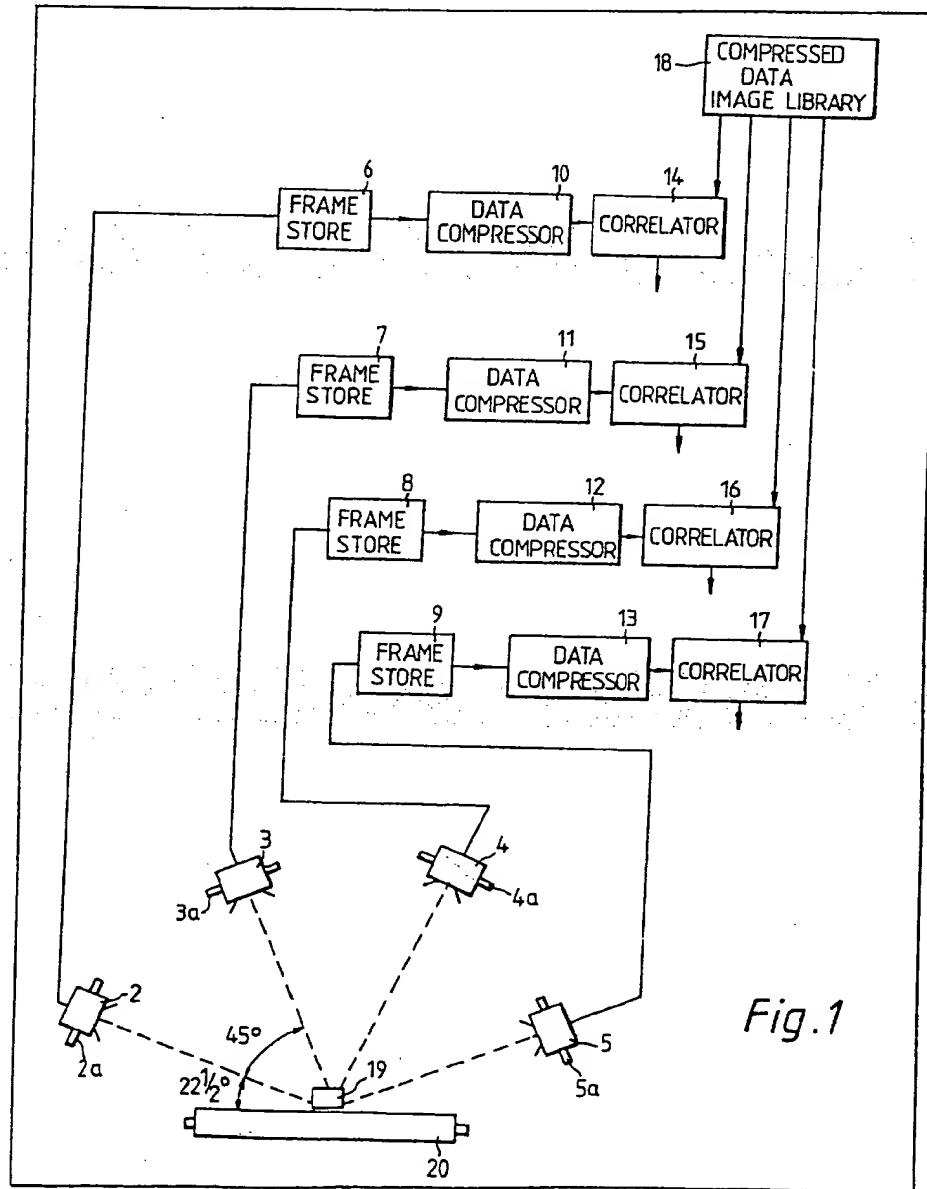
The drawings originally filed were informal and the print here reproduced is taken from a later filed formal copy.

2085629
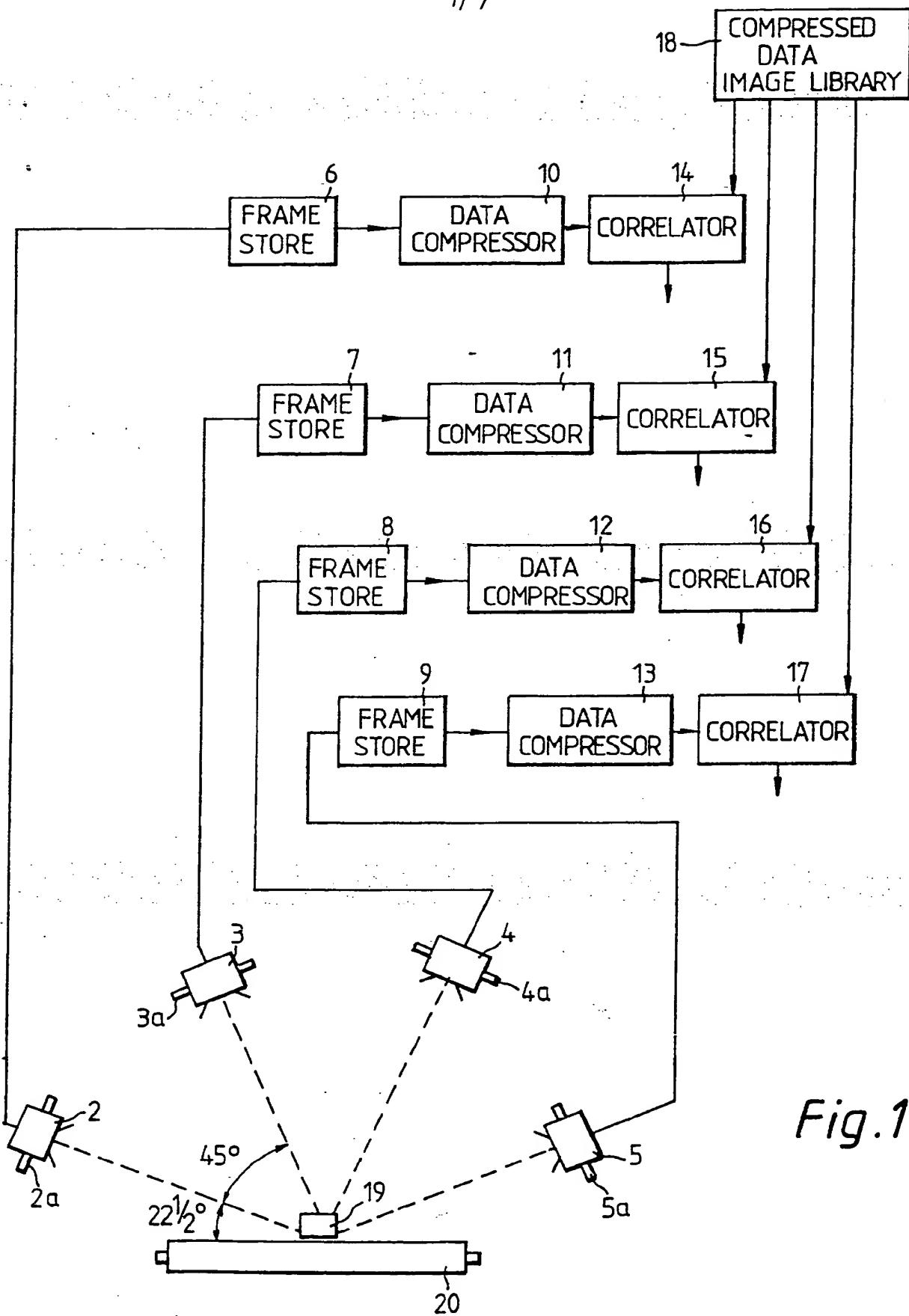
Fig.1
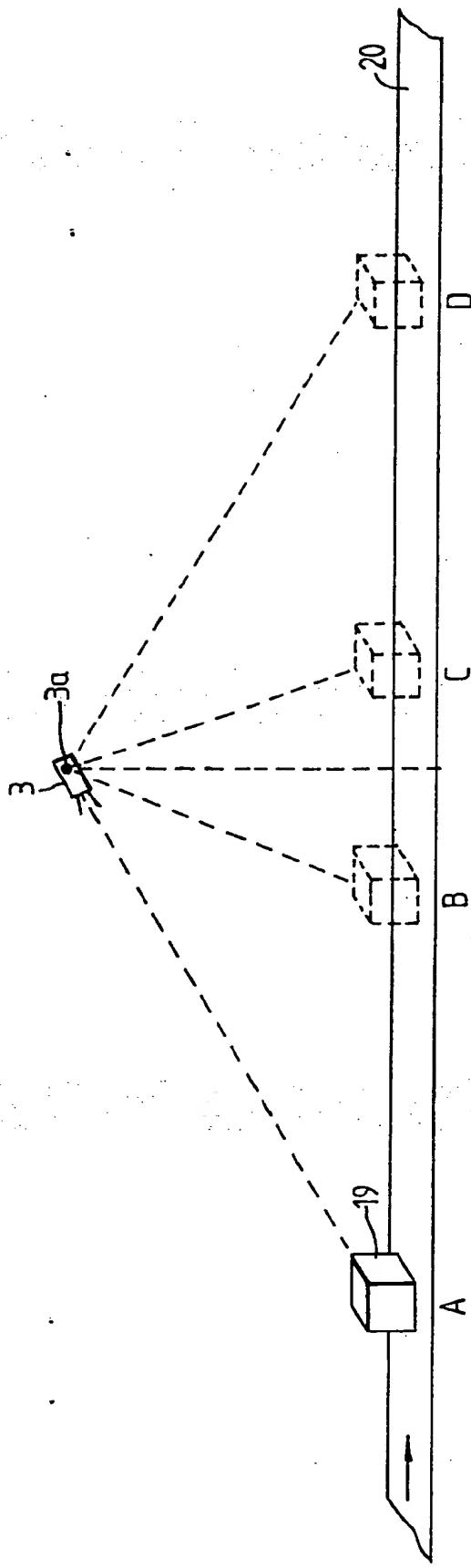
2085629

Fig.2
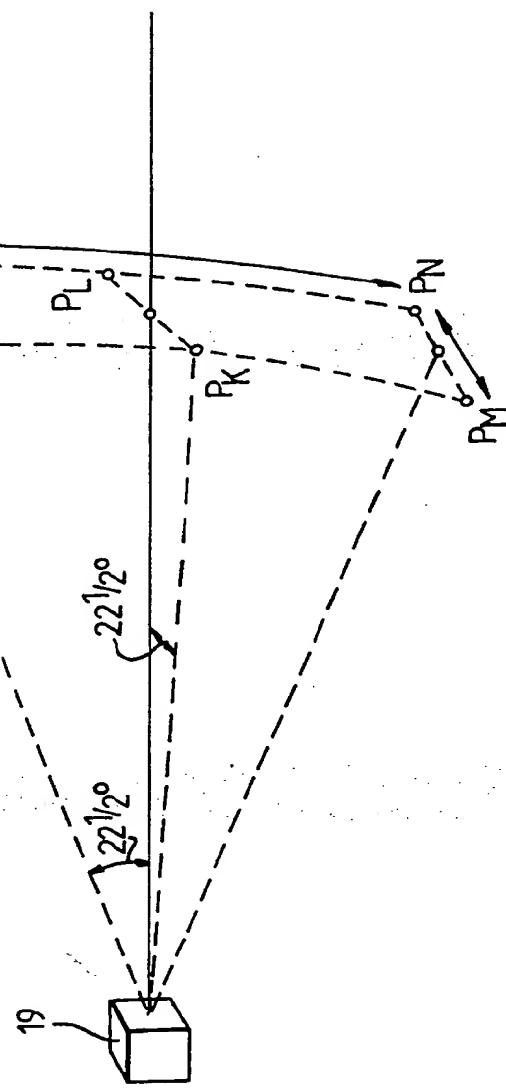
Fig.3
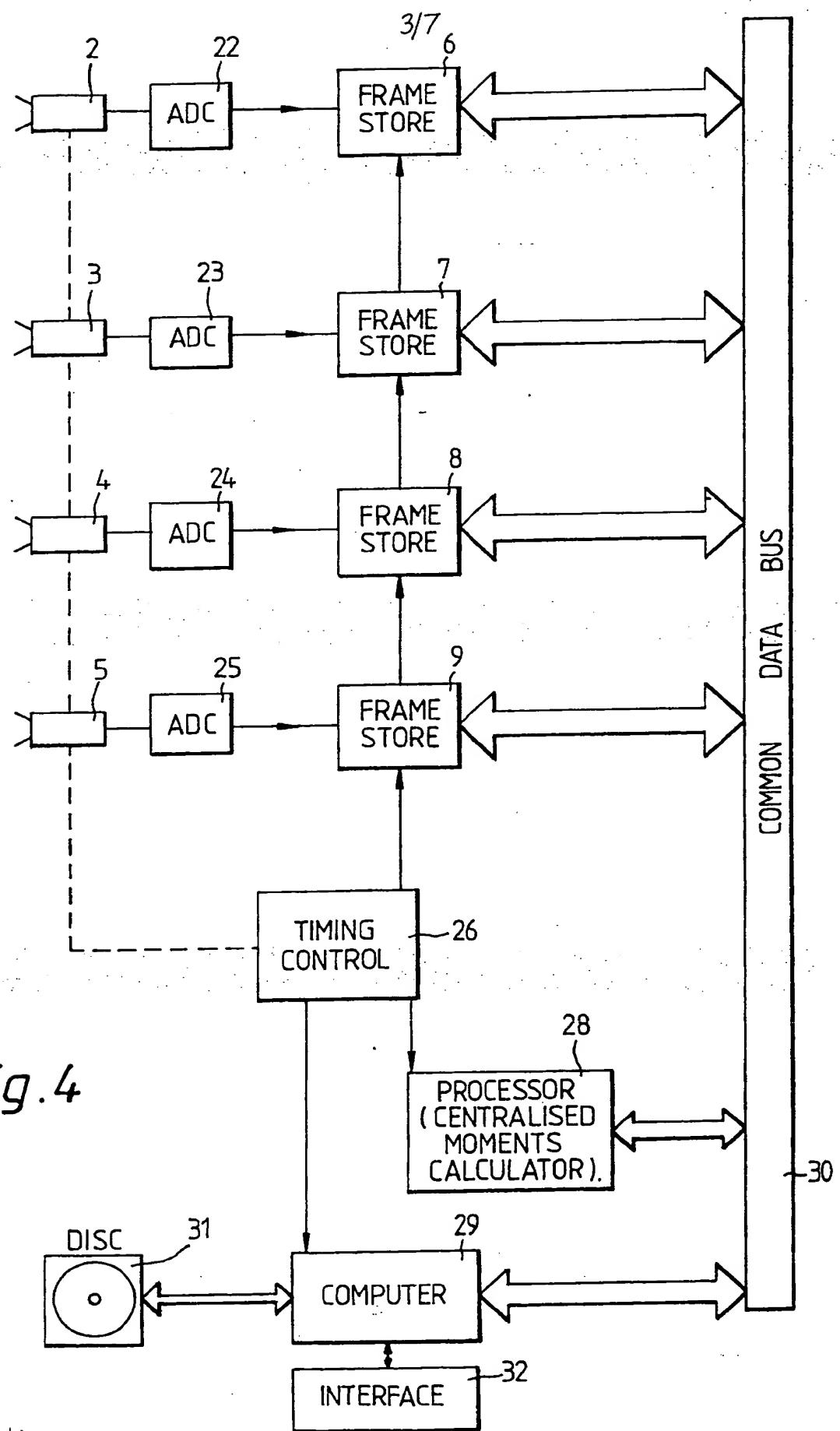
2085629

Fig.4
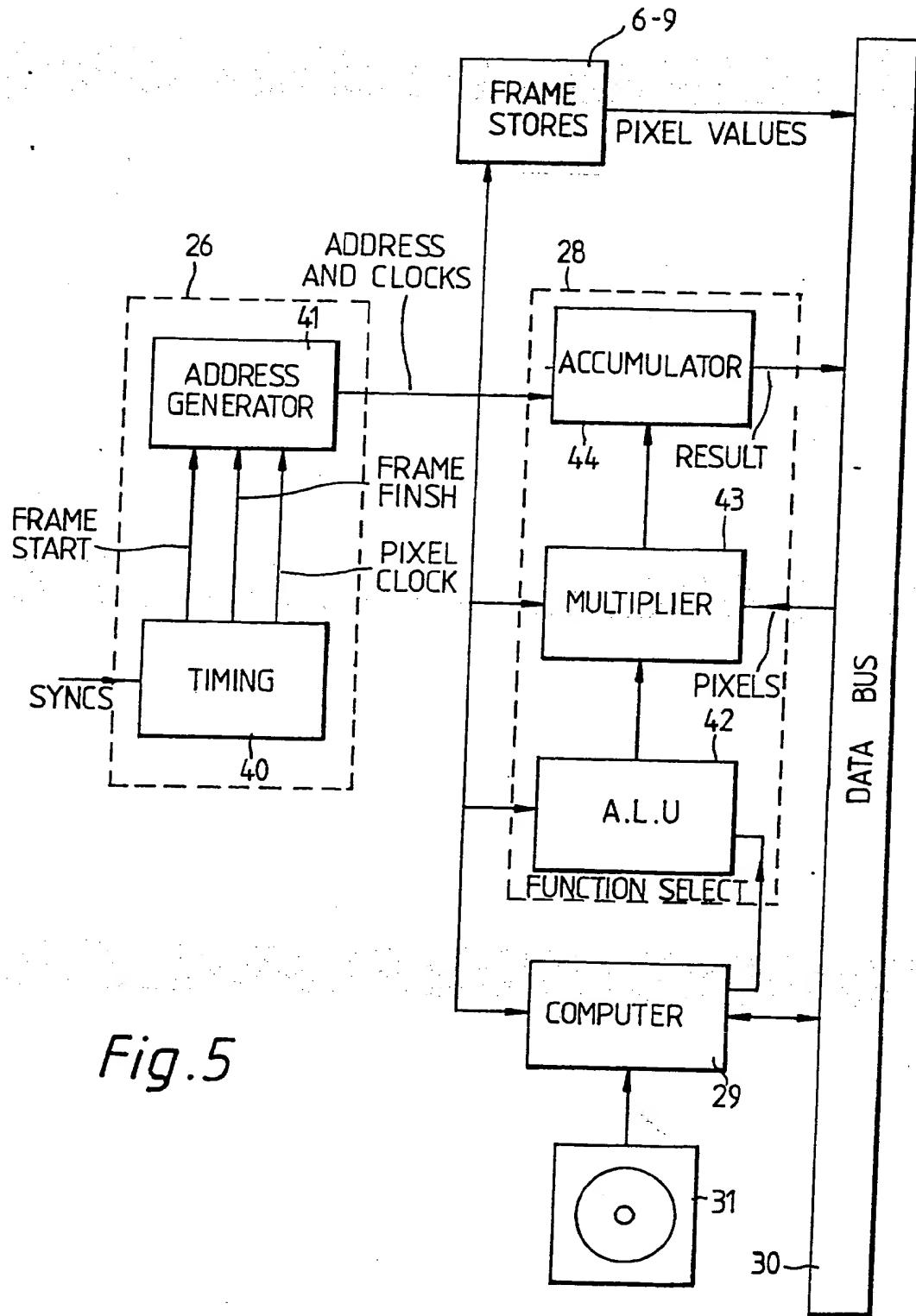
2085629

Fig.5

Fig.6

Logic path to calculate $\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y)$

Begin.

set sum1 = 0
sum2 = 0
sum3 = 0
sum4 = 0

for all x,y

set a=b=i=j=0
c=f(x,y)

apply TERM
calculation

add TERM
result to sum1

for all x,y

set a=x , i=1
b=j=0
c=f(x,y)

apply TERM
calculation

add TERM
result to sum2

for all x,y

set a=i=0
b=y, j=1
c=f(x,y)

apply TERM
calculation

add TERM
result to sum3

set $\bar{x}$=sum2/sum1
$\bar{y}$=sum3/sum1

for all x,y

set a=x-$\bar{x}$, i=p
b=y-$\bar{y}$, j=q
c=f(x,y)

apply TERM
calculation

add TERM
result to sum4

end

μ_{pq} = sum 4
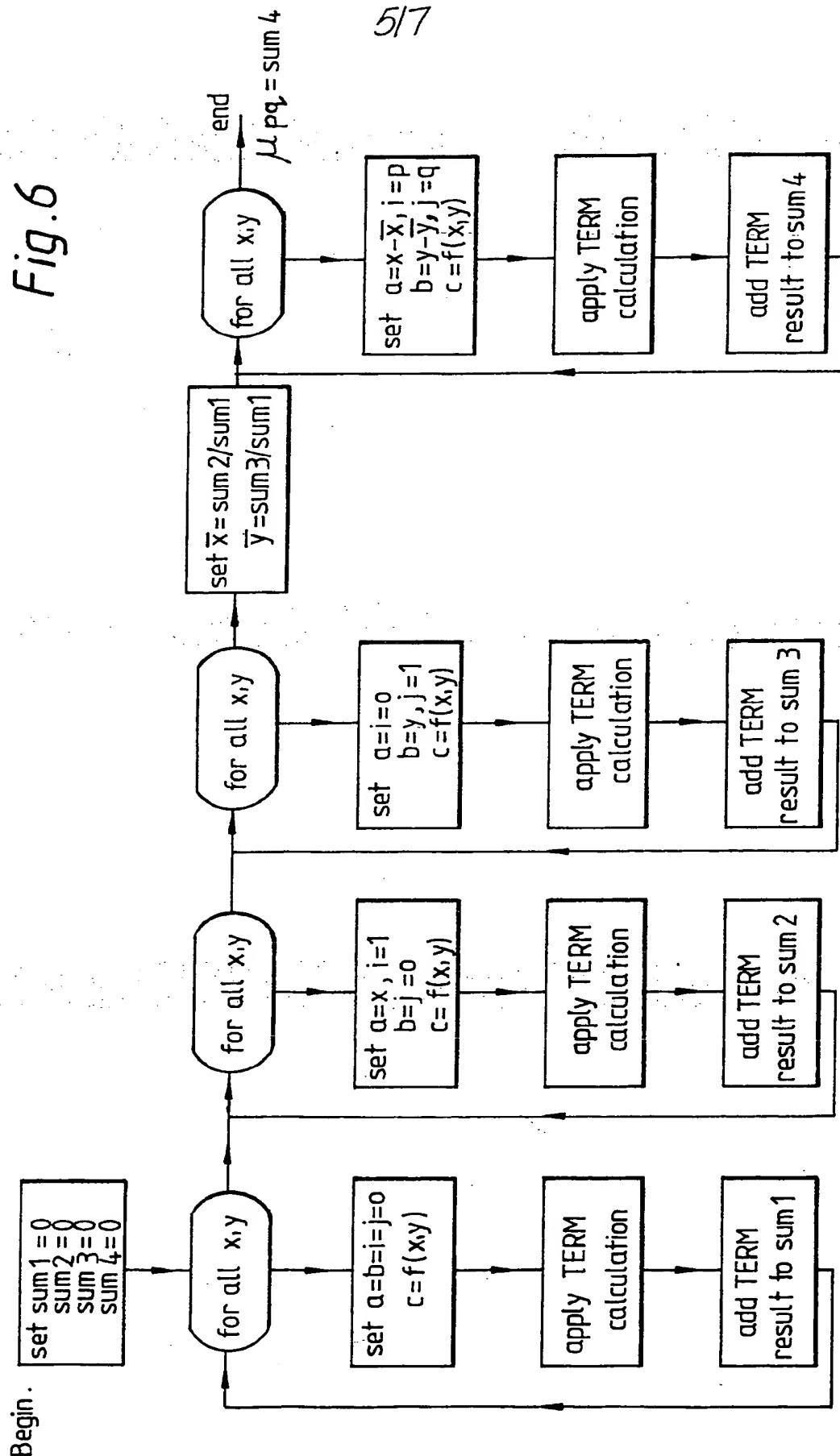
Logic path for 'TERM' calculation.

Input data values
a, b, c, i, j.

Fig.7

```
        ┌──────────────┐
        │  set TERM = 0│
        └──────────────┘
                │
                ▼
           ◇ if i = 0 ◇ ──────────────────────────┐
                │                                   │
                ▼                                   ▼
        ┌──────────────┐                      ◇ if j = 0 ◇ ──── output 'TERM' result
        │   multiply   │                           │
        │     TERM     │                           ▼
        │     by a     │                    ┌──────────────┐
        └──────────────┘                    │   multiply   │
                │                           │     TERM     │
                ▼                           │     by b     │
        ┌──────────────┐                    └──────────────┘
        │  subtract  1 │                           │
        │   from i     │                           ▼
        └──────────────┘                    ┌──────────────┐
                                            │  subtract  1 │
                                            │    from j    │
                                            └──────────────┘
```

2085629

Fig.8

FOUR IMAGES CAPTURED

CENTRAL MOMENTS CALCULATED

CORRELATED AGAINST STORED MOMENTS

TILT POSITION 4?

NO

YES

LINEAR PROJECTION OF SELECTED CAMERA IMAGE

ROTATE AND CORRELATE AGAINST MASTER
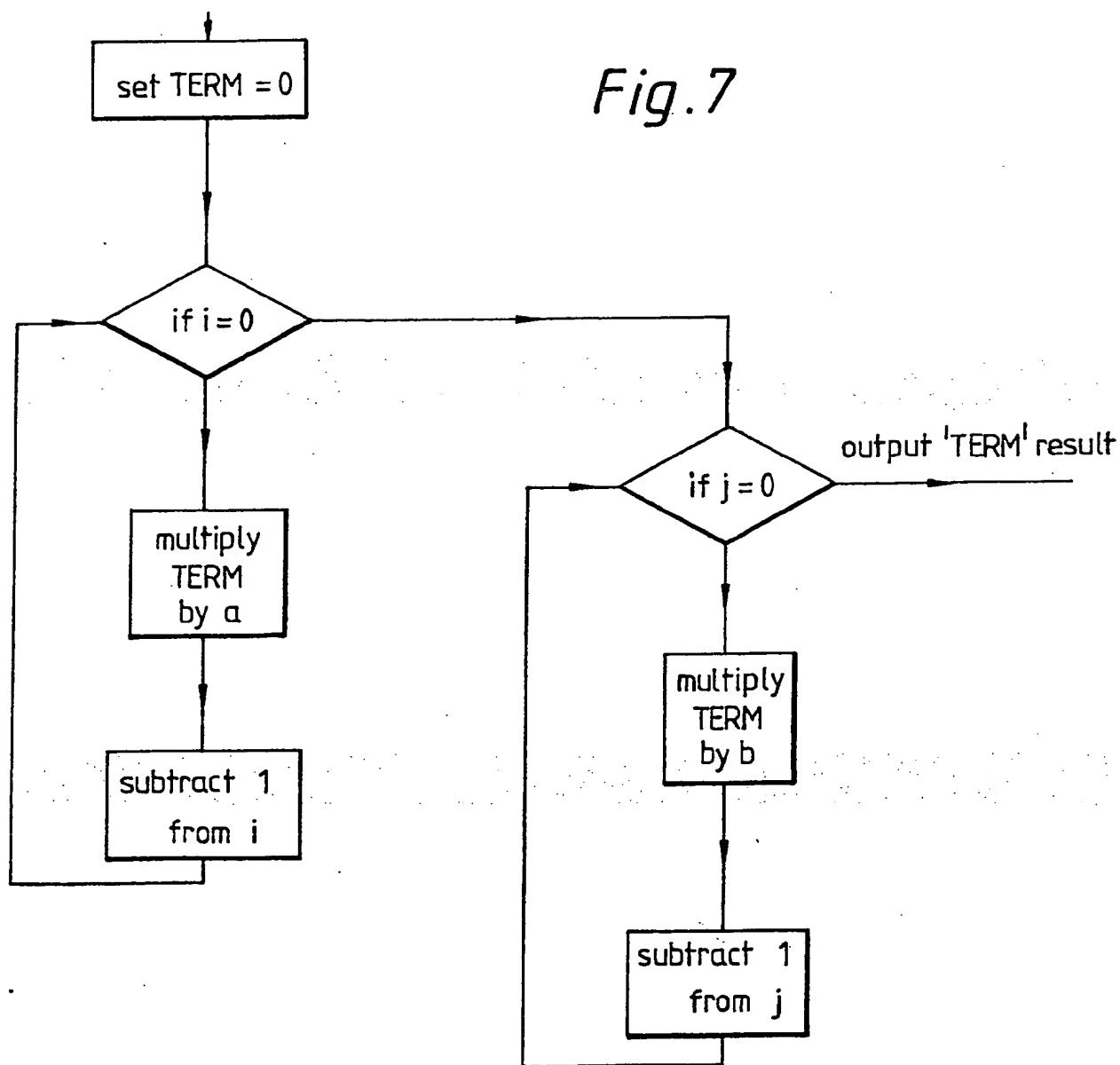
ORIENTATION DETERMINED.
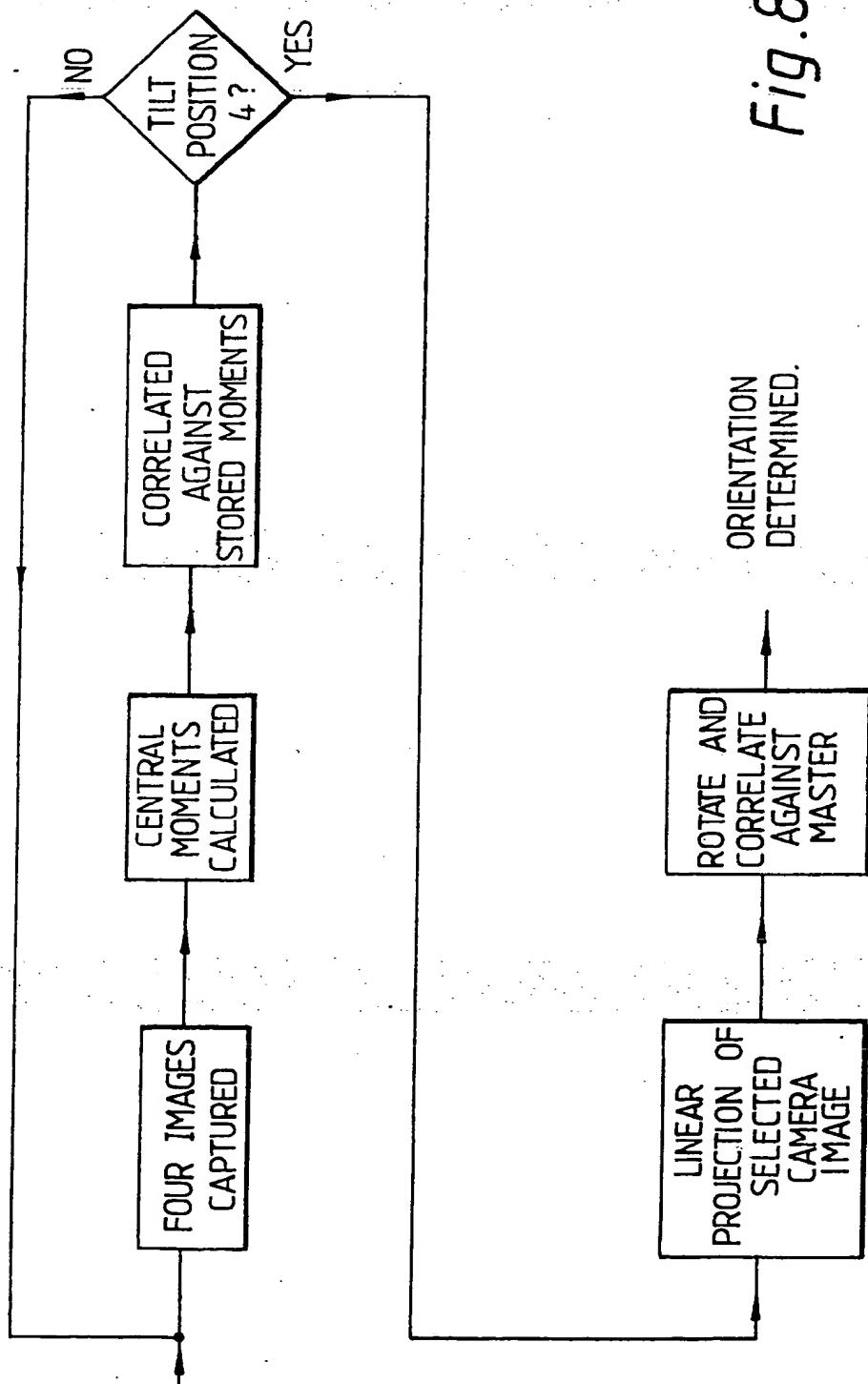
BLOCK DIAGRAM OF PROGRAM FLOW

SPECIFICATION
Object recognition

The invention relates to the recognition and/or determination of the orientation of an object, for example a machine casting moving on a conveyor belt.

5 The use of robots on the assembly process of industrial machinery is greatly curtailed by their lack 5 of vision. Even if vision is provided it usually only enables the robot to make a single simple decision. For example, it might be used just to confirm that a hole has been drilled into a part before assembly. In such known systems use is typically made of a single camera with the facility of comparing the information with a single stored image and it is required that the object be placed on a rotatable table

10 and by rotating this table, establish when correlation between the image from the camera and the 10 stored image is reached. The invention now to be described provides the robot or other interfaced device with far more powerful vision and thus greatly enhances its power of recognition.

According to the invention there is provided an object recognition system comprising means for capturing a plurality of images of the object each taken from a different aspect, processing means for

15 effecting data compression on the incoming image information to reduce the quantity of data presented 15 for recognition, correlation means for comparing the compressed image data with previously derived image data to determine whether any similarity is present.

The invention will now be described by way of example with reference to the accompanying drawings, in which:

20 FIGURE 1 shows the basic system of the present invention, 20
FIGURE 2 shows the selectable tilt positions of the cameras in use located over the conveyor belt,
FIGURE 3 shows possible camera positions used to compute the library of processed image data,
FIGURE 4 shows one arrangement for realising the invention,
FIGURE 5 shows aspects of the FIGURE 4 arrangement in more detail,

25 FIGURE 6 shows an example of the logic path employed to compute the centralised moments, 25
FIGURE 7 shows an example of the TERM calculation logic path, and
FIGURE 8 shows an example of a typical flow diagram of the system operation.

In industrial plants, machine parts or castings are often transported to the points of assembly by continuously moving conveyor belts or lines. A robot for example performing the assembly of parts must

30 be able to "see" the part's position and just as importantly, it's orientation so that it can pick up the part 30 in the correct way for assembly.

This visual capability is achieved in the present invention by processing images of the object under scrutiny taken from different aspects typically using a group of cameras and matching the processed data with previously manipulated stored image data of known orientation of that object as now

35 described. 35

The system of Figure 1 comprises four cameras 2—5, four frames stores 6—9, four data processors 10—13 and four correlators 14—17. The correlators have access to library 18. The object 19 of interest is shown on a conveyor 20 such that the object would move past the array of cameras (i.e. into the Z plane). Each of the cameras is provided to give a different viewing aspect of the object and the

40 output of the respective cameras is held in frame stores 6—9. The captured images are then passed to 40 respective data compressors 10—13 which provide an output derived from the image data but processed to simplify the correlation steps effected by respective correlators 10—13. The correlators have access to a library of images which have been previously processed in a similar way. The simplest way of producing such a library is to place the known object in various positions and store the

45 manipulated images from the compressors 10—13 directly in library 18. The cameras are shown in this 45 example as spaced apart at 45° relative to the object.

To increase the number of aspects the object is observed from the cameras 2—5 can be arranged to pivot about respective axles 2a—5a. This is especially necessary if the shape of the object is complex. Each camera could be tilted into four positions as shown in Figure 2 as the object (e.g. a casting) moves

50 along the belt. In the example the four positions A, B, C and D are illustrated to be at an angle of 45° 50 relative to the next or previous position. The cameras can be conveniently moved into the various tilt positions by being linked to the existing conveyor drive and making use of microswitches for example to determine when the desired tilt position has been reached.

Although Figure 1 shows a system with four cameras, the arrangement could be modified to use a

55 single camera to move in an arc and frame stores 6—9 would capture the image for a given position at 55 45° intervals. Similarly although the object is moving relative to the cameras, the system would operate if the cameras were operated to achieve the same relative movement.

When correlation at the output of a particular correlator 14—17 is achieved with the data from the library, the identity and orientation of the object will then be known.

60 The data compressors are required because the normal image typically of 256 × 256 picture 60 points would require correlation for each picture point and for each of the library of images until coincidence was detected. Such an operation would take too long to be practical and so it is necessary to manipulate the image data to provide simplified correlation yet retaining sufficient information to unambiguously identify the object observed.

IMAGE LIBRARY GENERATION

Prior to actual use of the system, a library of processed images of the casting must be generated. A direction of view of the casting should be chosen which will allow the 2 dimensional image as seen along this direction to uniquely determine the castings 3 dimensional orientation. Obviously rotation
5 about any axis of rotational symmetry can be ignored.

The view of the casting along the chosen line is taken by a single camera and captured with a typical resolution of 256 x 256 pixels by 4 bit greyscale.

The camera is then moved in angular increments in both the x and y directions whilst still pointing at the castings as shown in Figure 3. In this example the point P corresponds to the central viewing
10 position and the range of angular movement about the initial line of view is chosen to be +22½° to −22½° in both directions. Thus a number of images will be captured between the positions PK and PL and this approach is adopted throughout the viewing area bounded by the viewing positions PM to PQ. At each camera position within this range an image of the casting is processed to provide the compressed identity data. This information is typically stored in the library together with the angular
15 position of the camera.

The number of image positions actually used will determine the accuracy of the system. An array of 45 x 45 positions used to provide the library of image data will give a more than ½° error.

Thus the data compressor 10 provides an identification "fingerprint" for a given object orientation of sufficient detail to distinguish this from other object orientations yet allowing rapid identification from
20 the library. When identity is detected by a correlator, since the library position associated with this image is known and the original orientation and object is known, this data is sufficient to instruct any interfacing device which can then be caused to respond accordingly. Such data can be translated into machine instructions via a look up table for example.

It is to be appreciated that the object under scrutiny will typically be of more complicated shape
25 than the simple representation in the drawings.

One example for providing suitable compression techniques will now be described. In this example it is now assumed that the techniques employed for compression will make use of "Moment Invariants", although other suitable compression techniques could be adopted.

Thus processing of the image data to produce the simplified identification is mainly concerned
30 with calculating their normalized central moments. The calculation of such moments is known, see MING-KUEI HU "Visual Pattern Recognition by Moment Invariants" IRE Transactions on Information Theory (1962) Vol. IT—8 pp. 179—187. This technique is adopted in this system and expanded to cope with the three dimensional objects which require recognition. Such moments are invarient to translation, rotation and scale change and as such are an ideal means of matching a captured image of
35 the casting with a library of stored images. This first matching step results in the determination of one angle of orientation. A second step is required to obtain the other angle of orientation in a perpendicular plane.

An arrangement based on Figure 1 and suitable for handling the necessary processing is now shown in Figure 4. The four cameras 2—5 are shown as analog devices with outputs passed to the
40 digital frame stores 6—9 via associated analog to digital converters 22—25. Each of the frame stores is expediently shown having access to a common video data bus 30 with a computer 29 and a processor 28. Although a common system is shown, individual store control and processing may be utilised along the lines described in U.S. Patent 4,148,070 for example. Each framestore holds the pictures captured from the four positions of its corresponding camera. The capture of the image and its position within the
45 frame store is controlled by timing signals provided by control 26 in normal manner. The frame stores are all linked by the common data bus 30 to the purpose built hardware within processor 28 used to calculate the centralised moments. This processor provides the compression function represented by blocks 10—13 of Figure 1. Also on the data bus is the host computer 29 with access to a fast mass storage device acting as the image library and in this example is represented by magnetic disc 31.
50 The computer 29 and processor 28 are allowed access to the data when permitted by timing control 26 so that the centralised moments can be calculated (in processor 28) and the correlation effected (conveniently by computer 29). The computer can conveniently be programmed to provide the desired control function to the external machine or device via interface 32, rather than requiring the hardware look up table technique referred to above.
55 Starting with the cameras in the first position (position A of Figure 2), four images are captured by the 4 cameras. The video function processor 28 calculates the normalized central moments of each image. This can be typically achieved in less than one second using the hardware based system described below. Each image in turn is compared with the library of stored images by a simple correlation of the seven invarient moments. The highest correlated pair of images (one from that 4
60 camera view and one from the library) is recorded and updated as the library search is made. The time for the search and correlation depends on the number of images stored in the library, but even with the number in the thousands, the time is about one second. The whole process is repeated with the casting and cameras in the second, third and fourth positions (positions B—D).

The optimum correlated pair readily define within the resolution of the incremental angular
65 changes of the library images, the line chosen to uniquely determine the orientation of the casting.

A further step is required to obtain the orientation of the casting about this line. The 4 bit camera image which appeared in the optimum pair is projected onto the place perpendicular to the line. The image is repeatedly rotated (using the digital data) and correlated with the 4 bit stored library image using known techniques. The two images are not necessarily of the same scaling. At each rotation one image must be scaled so that the rectangle defined by the extremities of the image in the x and y direction is the same for both images. the optimum correlation provides the second angle of orientation of the casting.

As already explained with reference to Figure 3, a library of image data is built up for use in the recognition system. In this embodiment, the normalised central moment is calculated for each camera position and these are stored to give no more than $\frac{1}{2}°$ error.

NORMALIZED CENTRAL MOMENTS
For a digital image a central moment is defined as:

$$\mu pq = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x,y)$$

where $f(x,y)$ is the greyscale value at point (x,y)

$$\bar{x} = \sum_{x,y} x\, f(x,y) \left. \sum_{x,y} \right/ f(x,y)$$

$$\bar{y} = \sum_{x,y} y\, f(x,y) \left/ \sum_{x,y} f(x,y) \right.$$

and all summations are taken over the interior points of the image.
Normalized central moments are defined as:

$$\eta pq = \frac{\mu pq}{\mu oo^\gamma}$$

$$\text{where } \gamma = \frac{p+q}{2}$$

From the second and third normalized moments, a set of seven invarient moments can be derived (see Hu referred to above). These moments are invarient to translation rotation and scale change.

The processor 28 used to calculate the centralised moments will now be described by way of example with reference to Figure 5.

The timing block 40 within timing control 26 provides signals for initialisation and synchronisation of the processor hardware and the other system blocks. The address generator block 41 receiving frame timing and pixel clocks from block 40 provides the next frame store address whose pixel value is required in the moment calculation. This can be passed directly to the frame stores as shown or via the data bus 30 dependent on timing requirements. The ALU block 42 is a programmable arithmetic and logic unit which can calculate powers of the coordinates of the address as required by the moment definition. The function applied to the address by the ALU is selectable from the host computer 29 directly as shown or alternatively via the common data bus 30. The multiplier 43 allows the pixel value returned from the frame store together with the power of the address from the ALU to be multiplied. The accumulator 44 allows the result to be accumulated over an image and the result is passed to the host computer 29 to be correlated with stored values from disc 31.

As already explained the central moment is defined as

$$\mu pq = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x,y)$$

and thus some of the selected functions from the ALU 42 would typically be:

$$\mu 00 = \sum_x \sum_y f(x,y)$$

$$\mu 10 = \sum_x \sum_y (x - \bar{x}) f(x,y)$$

$$\mu 02 = \sum_x \sum_y (y - \bar{y})^2 f(x,y)$$

where f(x,y) is the pixel value at coordinate (x,y) in the frame store.

Although the system is shown employing ALU 42 dedicated hardware function blocks could alternatively be used.

Typical logic flow paths for the normalised central moments and "Term" calculations used therein are shown in Figures 6 and 7 respectively.

The present system using calculated moments enables the orientation of an object to be recognised by using only linear movement of the conveyor (without requiring the rotating table) and can handle various types of objects to say pick up one type without disturbing others.

An example of a typical application is shown in the flow diagram of Figure 8 which shows additional manipulation steps effected after the movement correlation step effected by the computer 29, employing standard techniques.

Thus the calculation for central moments is shown followed by correlation and this is repeated for all 4 tilt positions of the camera. After completion of this stage the linear projection of the selected image is rotated and correlated against the previously stored selected image to determine the orientation.

The objects recognised need not be restricted to those moving along a conveyor nor need the recognised data be used to control a robot exclusively. Other camera configurations could also be employed.

CLAIMS

1. An object recognition system comprising: means for capturing a plurality of images of the object each taken from a different aspect, processing means for effecting data compression on the incoming image information to reduce the quantity of data presented for recognition, correlation means for comparing the compressed image data with previously derived image data to determine whether any similarity is present.

2. A system as claimed in claim 1, wherein an image library device is provided containing a plurality of previously captured and processed images for use by the correlation means.

3. A system as claimed in claim 1, 2 or 3, wherein the means for capturing the images is adapted to provide aspects in more than one plane.

4. A system as claimed in claim 1, 2 or 3, wherein the means for capturing the images comprises a plurality of frame stores receiving data from at least one camera.

5. A system as claimed in claim 4, wherein the at least one camera is adapted to be pivoted to follow the movement of an object so as to generate the more than one image aspect.

6. A system as claimed in any one of claims 1 to 5, wherein the processing means is adapted to calculate the centralised moments of the captured object.

7. A system as claimed in claim 6, wherein the processing means includes a function controller and arithmetic device for calculating the centralised moments in steps determined by the function provided from said controller.

8. A system as claimed in any one of claims 1 to 7, wherein the correlation means includes an orientation manipulator to allow the relative orientation of the object to be determined.

9. A system as claimed in any one of claims 1 to 8, wherein a control interface is provided to allow the image identification to be used to effect control of an interfaced deviced connected thereto.

10. An object recognition system substantially as described herein and as illustrated in the accompanying drawings.